Quantitative proteomics: challenges and opportunities in basic and applied research

Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., & Aebersold, R. (2017). Quantitative proteomics: challenges and opportunities in basic and applied research. *Nature Protocols*, *12*(7), 1289-1294. https://doi.org/10.1038/nprot.2017.040

Published in:
Nature Protocols

Document Version:
Peer reviewed version

Quantitative proteomics: Challenges and opportunities in basic and applied research

Olga T. Schubert[1], Hannes L. Röst[2,5], Ben C. Collins[3,5], George Rosenberger[3,5,] Ruedi Aebersold[3,4,*]

1 Department of Human Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA

2 Department of Genetics, Stanford University, Stanford, CA 94305, USA

3 Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland

4 Faculty of Science, University of Zurich, 8057 Zurich, Switzerland

5 Authors contributed equally

*Corresponding author (aebersold@imsb.biol.ethz.ch)

Summary

In this perspective, we discuss developments in mass spectrometry-based proteomic technology in the last decade from the viewpoint of our laboratory. We also reflect on existing challenges and limitations, and explore the current and future role of quantitative proteomics in molecular systems biology, clinical research and personalized medicine.

Introduction

Proteins constitute a large part of the molecular machinery of the cell and are the major class of biomolecules targeted by drugs. Organized in functional modules and networks, they carry out cellular functions and determine phenotypes by means of coordinated activities of a multitude of molecular species[1]. Traditional biochemical methods for studying proteins have been highly biased towards a relatively small subset of proteins for which high quality, mainly antibody-based assays have been available[2]. Over the past two decades, mass spectrometry (MS)-based methods have emerged as the method of choice for the confident and near exhaustive identification and quantification of the proteins contained in a biological sample and have significantly contributed to unraveling cellular signaling networks, to elucidating the dynamics of protein-protein interactions in different cellular states, and to improved diagnosis and molecular understanding of disease mechanisms. Overall, MS-based proteomics can reveal the quantitative state of a proteome and thereby provides insights into the biochemical state of the respective cell or tissue. In the following paragraphs, we will discuss important concepts and developments in proteomic technology and explore the current and future role of quantitative proteomics in molecular systems biology as well as clinical research and personalized medicine.

Evolution of MS-based quantitative proteomics

MS-based proteomics can be broadly grouped into top-down proteomics where intact proteins are measured and bottom-up proteomics where peptides are measured as surrogates for the respective protein; in this commentary, we will focus on bottom-up proteomics. The typical bottom-up proteomics workflow starts with trypsin digestion of a protein sample into short peptides which are then separated by liquid chromatography either directly or after biochemical fractionation (Figure 1A)[3]. As peptides elute from the chromatography column, they are subjected to electrospray ionization[4,5] and are directly sprayed into the mass spectrometer, where two levels of MS measurement take place in tandem[3]. At the first level, a mass analyzer measures the mass-to-charge ratio (*m/z*) of peptide molecular ions (MS1). At the second level, *m/z* values of fragment ions resulting from the fragmentation of specific peptide ions are detected (MS2). The specific fragment ion pattern of each peptide ion together with its *m/z* value enable confident identification of peptides present in the sample. Identified peptide sequences can then be mapped to proteins and the signal intensities of either peptides or fragment ions can be used to estimate relative abundance changes across samples.

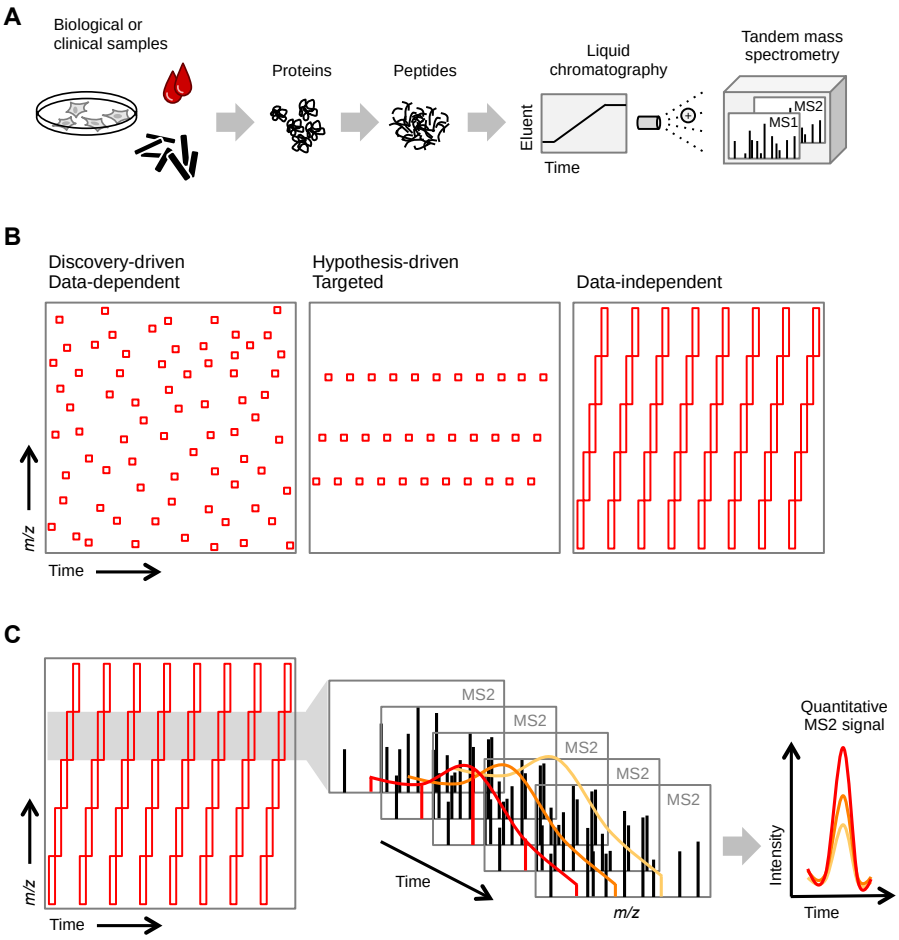

Figure 1. Standard MS-based proteomics workflow and acquisition schemes. (A) Proteins can be extracted from various biological samples, such as bacterial or mammalian cell culture, tissues or bodily fluids. They are then enzymatically digested into peptides, which are then subjected to reverse-phase liquid chromatography, ionized with electrospray ionization and sprayed into the mass spectrometer. The time dimension in B and C is this chromatographic time. (B) Different acquisition schemes for tandem MS sample the proteome in distinct ways. While the most widely used untargeted (also referred to as shotgun or data-dependent acquisition, DDA) is relatively simple and applicable to any sample without requiring prior knowledge, resulting data can suffer from missing data points due to the stochastic sampling process. In contrast, targeted acquisition acquires peptide and fragment ion data in a highly consistent manner allowing accurate and sensitive quantification, but is limited to a relatively small, pre-defined set of peptides. Data-independent acquisition (DIA) acquires data of all detectable fragment ions in a sample in a systematic and consistent manner, but due to the relatively large peptide ion isolation windows (*m/z* dimension) the resulting data is more complex than for the other two acquisition schemes. (C) DIA data can be analyzed in different ways, either directly analyzing the multiplexed MS2 spectra or first extracting a subset of informative fragment ion signals (requires prior knowledge) and using these to derive quantitative data for specific peptides[36,40].

To account for technical variability at various stages of sample handling and during the actual measurement, in the mid-90s, we and others started to develop strategies based on isotopic labeling[6,7], including chemical isotopic labeling[8], metabolic isotopic labeling[9], and isobaric tagging[10,11]. Another important application of isotopic labeling in MS is the use of labeled spike-in peptides or proteins of known concentration that enable the determination of absolute concentrations of proteins in a sample, for example, in terms of number of molecules per cell or nanograms per milliliter of blood[12]. While label-based approaches are still the gold-standard for quantification by MS-based proteomic methods[13], the past years have seen label-free approaches becoming more popular thanks to simpler experimental design and sample preparation[6,14]. Among the developments enabling this transition are the advance of commercially available high resolution and fast scanning instruments, such as the introduction of the Orbitrap (2005)[15] and continuous improvement of time-of-flight mass spectrometers[16], combined with improvements in software for aligning multiple MS runs[17,18]. Another more recent trend, starting in 2006[19], is label-free absolute quantification, where the absolute concentrations of all proteins measured in a sample are estimated based on summarized ion counts, which can then be converted into a meaningful unit by comparison to the total amount of protein that was injected into the mass spectrometer or by correlation to a set of spiked-in reference peptides of known concentration[20-23].

Regardless of whether label-based or label-free strategies are used, bottom-up proteomic methods have traditionally been divided into discovery proteomics and targeted proteomics (Figure 1B). Discovery proteomics (also known as shotgun proteomics and exemplified by data-dependent acquisition, DDA) has its strength in identifying thousands of proteins per run. However, in complex samples, we have often been faced by limitations regarding repeatability of peptide identification and consistency of quantification[24,25]. Recent developments in chromatographic performance and MS hardware alleviate some of these concerns and allow high-quality quantitative measurements of near-complete proteomes, even in highly complex samples such as human cell lines and tissues[26-29].

About a decade ago, in order to overcome the limited scalability and reproducibility of discovery proteomics in studies aiming to quantify proteins in cohorts consisting of large numbers of samples, we and others started exploring the capabilities of targeted proteomics (exemplified by selected/multiple reaction monitoring, S/MRM[30,31], and more recently parallel reaction monitoring, PRM[32,33]). Targeting methods provide consistent and accurate quantification, even

at low abundances and in complex mixtures. While targeted proteomics is typically limited to a few dozen predefined proteins per run, its sensitivity and highly quantitative capabilities make it well-suited for hypothesis-driven research and clinical studies where a smaller number of proteins, such as potential biomarkers, are to be measured in large numbers of patient samples[34].

Placed between these two well established techniques, a third type of mass spectrometric acquisition has gained remarkable momentum in the past five years: data-independent acquisition (DIA)[35,36]. In this method, multiplexed fragment ion spectra are acquired systematically using deterministic peptide ion isolation windows that collectively span the mass range in which most tryptic peptides are expected (Figure 1B). DIA was first described in the early 2000s[37,38] and the following years have seen various further implementations and developments of the concept[35,36]. While these methods were of substantial conceptual interest and also led to a commercial implementation referred to as MS^{E} [39], adoption of them in the field was somewhat limited due to the overwhelmingly complex data resulting from their application to high-complexity samples. In 2012, our lab described a new DIA-based method termed SWATH-MS, which uses a targeted paradigm for the analysis of DIA data[40]. The novel analysis strategy based on comprehensive spectral libraries[41] and refined targeted scoring algorithms[42], together with improved instrumentation and an optimized acquisition scheme, enabled us to efficiently deconvolute the highly multiplexed DIA data and use it to achieve highly consistent quantification of thousands of analytes (Figure 1C). Latest developments to improve DIA methods include: coupling with ion mobility[43], new acquisition schemes[44], as well as new data analysis modes and software tools[45,46]. Combining the analyte throughput of discovery proteomics with the accuracy and repeatability of targeted proteomic methods, DIA/SWATH-type techniques have been applied successfully in a variety of studies and are becoming increasingly prevalent in the quantitative proteomics field, particularly in studies that require the consistent analysis of large sample cohorts[23,47-50].

Over the years, sensitivity and speed of mass spectrometers has improved at a rapid pace, however, obtaining robust, quantitative data over large numbers of samples remains one of the greatest challenges in proteomics, even for expert labs[51]. The most rigorous way to assess the capabilities and pitfalls of proteomic methods are inter-laboratory comparison studies. Such studies were conducted for discovery proteomics[24,52] as well as targeted proteomics[53]. In a recent study using DIA/SWATH-type methods[54,55], we found that over 4000 proteins from over

200 measurements of a human cell line can be independently identified and quantified across laboratories and instruments with coefficients of variation typically around 20%[54]. These results are encouraging because they show that quantitative proteomics is capable of delivering accurate, reproducible and comprehensive data at high throughput.

In the following two sections, we briefly outline our view of the fields' current and future role in molecular systems biology as well as in clinical research and applications.

Quantitative proteomics in molecular systems biology and the study of cellular organization

The comprehensive quantification of proteins, and their post-translational modification status across conditions, or over time in response to a stimulus or perturbation, is an important aspect of systems biological studies. Thanks to the technological advances described above, the quality of the resulting quantitative data matrices for large sample numbers has substantially improved, enabling us and others to conduct systems-oriented studies, not only in microorganisms, but also higher organisms, including mammals[23,47,50,56,57].

Proteins are not isolated molecules but three-dimensional objects acting in the context of other proteins, the modular and spatial organization of proteins can therefore be as important as their expression levels[1,58]. MS-based proteomic methods developed to query the organizational units of the proteome typically combine MS measurements with biochemical assays (Figure 2). The oldest of these methods, first described in 1999[59], is affinity purification coupled to MS to find interaction partners of a specific protein[60-62]. More recently, proteome fractionation using native separations has been applied to study protein complexes in a cell on a proteome-wide scale[63-65]. To determine subunit topologies of protein complexes and thereby obtain insights into the architecture of macromolecular assemblies, we and others have used approaches based on chemical cross-linking of protein residues[66-68]; adding a quantitative dimension enables probing of dynamic changes in protein complex composition and structure[69,70]. Furthermore, spatial resolution of the proteome within a cell can be added by combining proteomic techniques with enzymatic activities to label proximate or interacting proteins of a particular protein of interest[71]. Overall, these techniques highlight the power and flexibility of MS-based proteomics to not only produce a comprehensive and high-quality data matrix of protein abundances across a large number of samples, but also to obtain a dynamic, three-dimensional view of the proteome and

its modular and spatial organization, both of which are critical to fully understand complex biological systems.

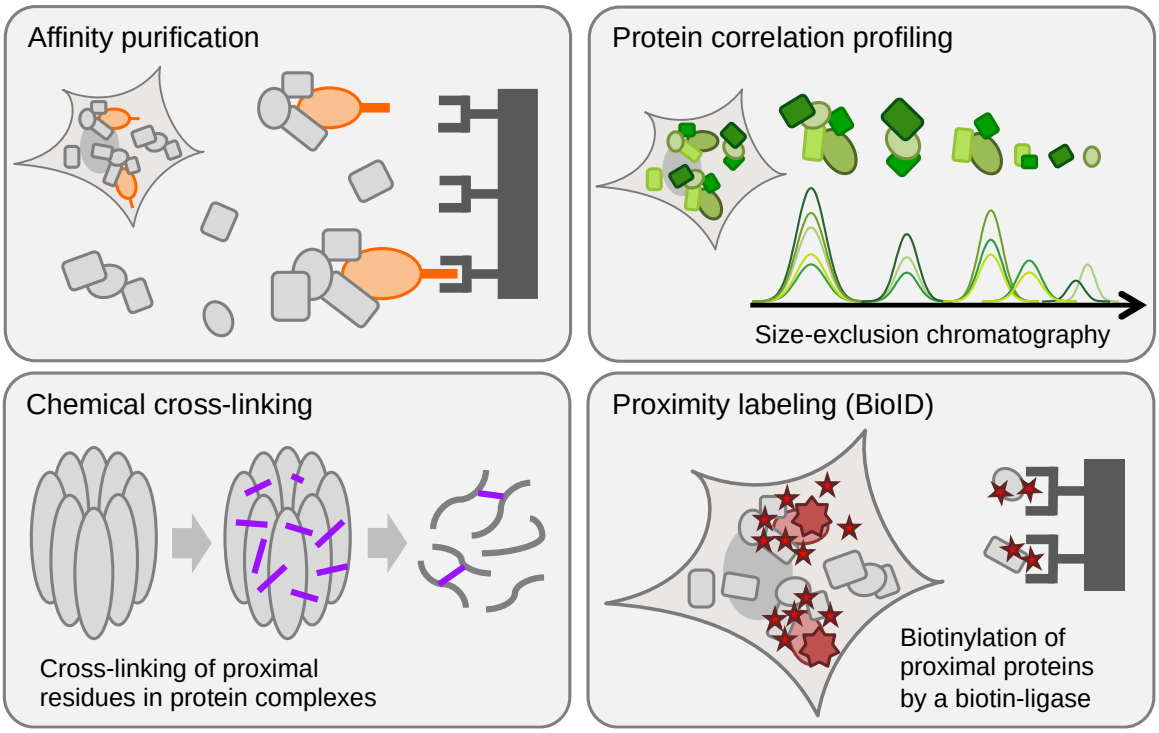

Figure 2. A selection of methods to explore the modular and spatial organization of the proteome. Two methods to study protein-protein interactions and complexes as they occur within a cell are affinity purification and protein correlation profiling. For both methods, cells are lysed under native conditions preserving protein-protein interactions. For affinity purification, a protein of interest is purified either via an affinity tag genetically added to the protein or with a specific antibody and subjected to MS analysis to identify proteins directly or indirectly binding to the protein of interest[60]. For protein correlation profiling, a cell lysate is fractionated e.g. by size-exclusion chromatography and after MS analysis of all the resulting fractions, correlation analysis is performed to find co-elution of proteins indicative of them having been part of the same protein complex[63-65]. Chemical cross-linking can be used to gain insights into the topology of a protein complex [66-68]. After digestion of cross-linked proteins, cross-linked peptides can be identified by MS and provide information on which parts of which proteins are in close proximity within the protein complex. Another emerging method is proximity labeling (also called BioID) using a ubiquitous biotin-ligase fused to a protein of interest to biotinylate all proteins in its proximity[71]. Biotinylated proteins can then be isolated and identified by MS. The BioID method captures not only stable protein complexes but also transient interactions between proteins that could not be captured by the other methods mentioned here.

Quantitative proteomics in molecular medicine

One of the major challenges in clinical studies is the requirement for large patient cohorts to cope with biological and experimental variability of clinical samples. In contrast to genomics, proteomic analyses of cohorts consisting of hundreds of samples are still prohibitively time-consuming and expensive, especially if large numbers of proteins are to be consistently quantified across the cohort. Two remarkable studies from 2016 used discovery proteomics to quantify several thousand proteins across over a hundred patient samples each; both efforts required several months of instrument time[72,73]. To conduct studies at larger scale, proteomic techniques that allow higher throughput, while maintaining robustness, repeatability and sensitivity are therefore essential. DIA/SWATH-type approaches emerge as a promising alternative for the quantitative proteomic analysis of clinical samples and first studies applied them successfully to quantify large numbers of proteins across hundreds of human patient samples[74].

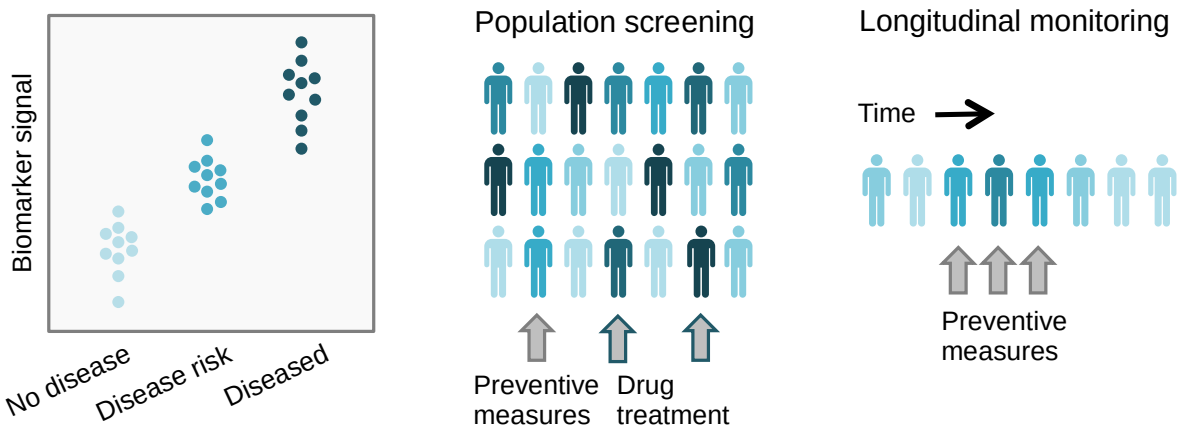

Figure 3. Quantitative proteomics in molecular medicine. Applications of quantitative proteomics in personalized medicine are typically based on biomarkers that reflect disease risk or disease status. Biomarkers are screened across individuals or patient cohorts. Longitudinal profiling of individuals allows monitoring of the molecular profile of a person over long time frames; and more meaningful clinical information can be obtained by comparing each measurement with previous time points of the same person than by comparing a single measurement with the population average. This personalized approach to molecular medicine is expected to achieve early and highly sensitive detection of disease risk and is therefore most effective in preventing disease.

The datasets generated from such studies can be used to discover cellular mechanisms and processes that are affected in the disease under study. Alternatively, quantitative proteomic techniques can be used to profile potential protein biomarkers in patient tissue, blood or urine to

inform disease risk, diagnosis, prognosis, and treatment stratification (Figure 3). Biomarkers also play a crucial role in the emerging field of personalized medicine, where recurrent molecular measurements of specific protein and metabolite levels are used to evaluate an individual's health status and to prevent development of disease in a timely manner by dietary, exercise or drug-based interventions (Figure 3)[75]. While protein biomarkers have traditionally been measured with immunoassays, targeted proteomic techniques have a number of advantages, including faster assay development, multiplexing capabilities and analytical specificity, and are therefore the method of choice to test panels of candidate biomarkers before they enter clinical validation studies[76-78].

Conclusion

Over the past two decades, we have witnessed rapid developments in mass spectrometric instrumentation as well as acquisition methods and analysis strategies. Furthermore, quantitative proteomics has contributed enormously to biological and clinically oriented research. However, current instrument operation as well as data acquisition and analysis still require highly specialized expertise. Many facilities, including ours, are therefore working towards the development of more robust MS-based methods and automated analysis pipelines to make quantitative proteomics available, not just to expert labs, but also to general molecular biology laboratories in academia, hospitals and industry.

Competing financial interest

RA holds shares of Biognosys AG, which operates in the field covered by the article. The remaining authors declare no competing financial interest.

Acknowledgements

OTS is supported by the Human Frontier Science Program (LT000737/2016) and the Swiss National Science Foundation (P2EZP3_165280), HLR is supported by the Swiss National Science Foundation (P2EZP3_162268) and EMBO (ALTF_854-2015), BCC is supported by a Swiss National Science Foundation Ambizione grant (PZ00P3_161435), and RA is supported by the Swiss National Science Foundation, the Swiss Initiative for Systems Biology (SystemsX.ch), and the European Union via ERC (AdG-670821).

Contributions

OTS, HLR, BCC, GR, and RA prepared the manuscript.

References

1. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402,** C47–52 (1999).
2. Edwards, A., Isserlin, R., Bader, G. & Frye, S. Too many roads not taken. *Nature* (2011).
3. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
4. Yamashita, M. & Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem* **88,** 4451–4459 (1984).
5. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246,** 64–71 (1989).
6. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Küster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389,** 1017–1031 (2007).
7. Bantscheff, M., Lemeer, S., Savitski, M. M. & Küster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404,** 939–965 (2012).
8. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19,** 1720–1730 (1999).
9. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1,** 376–386 (2002).
10. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75,** 1895–1904 (2003).
11. Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3,** 1154–1169 (2004).
12. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* **100,** 6940–6945 (2003).
13. Tabb, D. L. *et al.* Reproducibility of Differential Proteomic Technologies in CPTAC Fractionated Xenografts. *J Proteome Res* **15,** 691–706 (2016).
14. Chelius, D. & Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **1,** 317–323 (2002).
15. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40,** 430–443 (2005).
16. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M. & Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal Chem* **83,** 5442–5446 (2011).
17. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26,** 1367–1372 (2008).
18. Weisser, H. *et al.* An automated pipeline for high-throughput label-free quantitative proteomics. *J Proteome Res* (2013). doi:10.1021/pr300992u
19. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5,** 144–156 (2006).

20. Ahrné, E., Molzahn, L., Glatter, T. & Schmidt, A. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* **13,** 2567–2578 (2013).
21. Ludwig, C., Claassen, M., Schmidt, A. & Aebersold, R. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol Cell Proteomics* **11,** M111.013987–M111.013987 (2012).
22. Rosenberger, G., Ludwig, C., Röst, H. L., Aebersold, R. & Malmström, L. aLFQ: An R-package for estimating absolute protein quantities from label-free LC-MS/MS proteomics data. *Bioinformatics (Oxford, England)* **30,** 2511–2513 (2014).
23. Schubert, O. T. *et al.* Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of Mycobacterium tuberculosis. *Cell Host Microbe* **18,** 96–108 (2015).
24. Tabb, D. L. *et al.* Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* **9,** 761–776 (2010).
25. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* **28,** 710–721 (2010).
26. Gholami, A. M. *et al.* Global Proteome Analysis of the NCI-60 Cell Line Panel. *Cell Reports* **4,** 609–620 (2013).
27. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* **7,** 383–385 (2010).
28. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11,** M111.014050–M111.014050 (2012).
29. Wiśniewski, J. R. *et al.* Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Molecular Systems Biology* **8,** 611 (2012).
30. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138,** 795–806 (2009).
31. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **9,** 555–566 (2012).
32. Gallien, S. *et al.* Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell Proteomics* **11,** 1709–1723 (2012).
33. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **11,** 1475–1488 (2012).
34. Whiteaker, J. R. *et al.* A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* **29,** 625–634 (2011).
35. Chapman, J. D., Goodlett, D. R. & Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom Rev* **33,** 452–470 (2013).
36. Gillet, L. C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem* **9,** annurev–anchem–071015–041535–472 (2016).
37. Purvine, S., Eppel, J.-T., Yi, E. C. & Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3,** 847–850 (2003).
38. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R., III. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* **1,** 39–45 (2004).
39. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **77,** 2187–2200 (2005).
40. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11,** O111.016717–O111.016717 (2012).

41. Schubert, O. T. *et al.* Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat Protoc* **10,** 426–441 (2015).
42. Röst, H. L. *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **32,** 219–223 (2014).
43. Geromanos, S. J., Hughes, C., Ciavarini, S., Vissers, J. P. C. & Langridge, J. I. Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. *Anal Bioanal Chem* **404,** 1127–1139 (2012).
44. Egertson, J. D. *et al.* Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* **10,** 744–746 (2013).
45. Tsou, C.-C. *et al.* diA-umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* **12,** 1–14 (2015).
46. Wang, J. *et al.* MSPLIT-DIA: sensitive peptide identification for data-independent acquisition. *Nat Methods* **12,** 1106–1108 (2015).
47. Okada, H., Ebhardt, H. A., Vonesch, S. C., Aebersold, R. & Hafen, E. Proteome-wide association studies identify biochemical modules associated with a wing-size phenotype in Drosophila melanogaster. *Nat Commun* **7,** 12649 (2016).
48. Bruderer, R. *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14,** 1400–1410 (2015).
49. Selevsek, N. *et al.* Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-MS. *Mol Cell Proteomics* **14,** 739–749 (2015).
50. Williams, E. G. *et al.* Systems proteomics of liver mitochondria function. *Science* **352,** aad0189–aad0189 (2016).
51. Röst, H. L., Malmström, L. & Aebersold, R. Reproducible quantitative proteotype data matrices for systems biology. *Mol Biol Cell* **26,** 3926–3931 (2015).
52. Rudnick, P. A. *et al.* Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. *Mol Cell Proteomics* **9,** 225–241 (2010).
53. Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **27,** 633–641 (2009).
54. Collins, B. C. *et al.* Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *bioRxiv* 074567 (2016). doi:10.1101/074567
55. Navarro, P. *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. *Nat Biotechnol* 1–11 (2016). doi:10.1038/nbt.3685
56. Picotti, P. *et al.* A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494,** 266–270 (2013).
57. Wu, Y. *et al.* Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* **158,** 1415–1430 (2014).
58. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92,** 291–294 (1998).
59. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17,** 1030–1032 (1999).
60. Gingras, A.-C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8,** 645–654 (2007).
61. Collins, B. C. *et al.* Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat Methods* (2013). doi:10.1038/nmeth.2703
62. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162,** 425–440 (2015).
63. Foster, L. J. *et al.* A mammalian organelle map by protein correlation profiling. *Cell* **125,**

187–199 (2006).
64. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nat Methods* **9,** 907–909 (2012).
65. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150,** 1068–1081 (2012).
66. Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci USA* **97,** 5802–5806 (2000).
67. Rappsilber, J., Siniossoglou, S., Hurt, E. C. & Mann, M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* **72,** 267–275 (2000).
68. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc* **9,** 120–137 (2014).
69. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159,** 1042–1055 (2014).
70. Walzthoeni, T. *et al.* xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat Methods* **12,** 1185–1190 (2015).
71. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196,** 801–810 (2012).
72. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534,** 55–62 (2016).
73. Zhang, H. *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166,** 755–765 (2016).
74. Liu, Y. *et al.* Quantitative variability of 342 plasma proteins in a human twin population. *Molecular Systems Biology* **11,** 786 (2015).
75. Topol, E. J. Individualized medicine from prewomb to tomb. *Cell* **157,** 241–253 (2014).
76. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* **24,** 971–983 (2006).
77. Cima, I. *et al.* Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proc Natl Acad Sci USA* **108,** 3342–3347 (2011).
78. Surinova, S. *et al.* Non-invasive prognostic protein biomarker signatures associated with colorectal cancer. *EMBO Mol Med* **7,** 1153–1165 (2015).